

Transparency and explainability in data

Course Overview

Summary

As AI increasingly permeates our daily lives, from healthcare diagnoses to financial predictions, the need for transparency becomes paramount. Transparency not only builds trust but also serves as a cornerstone for accountability, responsible innovation, and ethical data practices. It provides insights into how AI systems operate, how they handle data, and how they navigate the complexities of decision-making. Just as we seek fairness in the results generated by AI, we also strive for clarity in understanding the systems that produce those results.

This three-part course will illuminate the multifaceted world of transparency and explainability within data and AI systems. We will begin by exploring the transparency spectrum, examining what transparency truly means, and understanding its benefits and challenges.

In the second section, we will delve into the various approaches organisations can employ to enhance transparency in their data and AI systems. From technical transparency to process transparency and outcome transparency, we will uncover the strategies that empower stakeholders to gain a deeper understanding of these complex systems.

Lastly, we will venture into the realm of explainable data and AI (XAI). As AI systems continue to make impactful decisions, it's imperative that we can interpret and trust their processes. This section will introduce the benefits of explainability, explore techniques to make data and AI explainable, and address the current challenges and approaches in this critical area.

By the end of this course, you will not only appreciate the significance of transparency and explainability in data and AI systems but also possess the knowledge to navigate and advocate for these essential principles more widely.

Learning Outcomes

In this course focused on transparency and explainability, you will explore key aspects of building trust in data practices and AI models, developing your understanding and the approaches to building trust.

To achieve this, you will:

- Understand the importance of transparency in data practices, and effectively communicate the significance of transparency in data practices, emphasising its role in building trust
- Implement techniques for developing AI models that are explainable and interpretable, enabling a better understanding of model decisions
- Create transparency measures to build trust with stakeholders, demonstrating how openness and clear communication contribute to ethical and responsible data and AI practices.

Learning Experience

Number of modules	3
Modality	Asynchronous / Self-directed / Online
Notional learning hours	2 hours (total)
Assessment	Formative
Certificate	Certificate of completion

Module Summary

Module Name	Description
The Transparency Spectrum	<p>AI is significantly impacting our lives. There are increasing numbers of data-enabled applications used across industries, from healthcare and finance to education and criminal justice. With this growing ubiquity, transparency is the key to building trust, promoting fairness, enabling accountability, and ensuring responsible innovation. It supports stakeholders in understanding how a system works, including how it handles data, its limitations and potential biases, as well as its context of use.</p> <p>This module will cover:</p> <ul style="list-style-type: none"> ● What is transparency ● Benefits and challenges of transparency ● The transparency spectrum ● Ethical Guidelines for transparency
Transparency approaches for data and AI	<p>This module introduces the three primary approaches that organisations can adopt to foster transparency: technical transparency, process transparency, and outcome transparency.</p> <p>It also provides practical examples, real-world use cases, and essential metrics to measure the effectiveness of these transparency efforts.</p> <p>This module will cover:</p> <ul style="list-style-type: none"> ● Transparency approaches ● Practice examples and case studies in each approach ● Transparency metrics and performance indicators ● The relationship between transparency and trust

Explainability and interpretability of data and AI systems

With the increasing integration of AI technologies into decision-making processes, there is common discourse in research and policy circles regarding the degree to which individuals can understand how these systems work, whether they are interacting with AI, or affected by AI-driven decisions.

Explainable data and AI (XAI) offers the potential to uncover how AI technologies work by building trust, enabling interpretation of the decision-making processes, and ensuring organisations are responsible in their use of data and development of models.

This module will cover:

- Benefits of explainable data and AI
- Techniques to make data and AI explainable
- Current challenges to explainability
- Approaches to explainability